# Linking Vegetation Productivity, Climate and Grazing Activities in Southeast Arizona

Guillermo Ponce, Phil Heilman, Gerardo Armendariz, Eric Moser, Vince Archer, Robert Vaughan

RISE
October 17th, 2015



# On the importance of AZ Grazing Activities

- Grazing is the dominant land use in Arizona
- Grazing land makes up 73 percent of Arizona's total land area
- Approximately 98 percent of Arizona's total agricultural land
- Cropland accounting for the remaining 2 percent of agricultural land

*Source:BLM*

Source: *The Contribution of the Beef Industry to the Arizona Economy (*Kerna et al., 2014).

BLM

USDA FOREST SERVICE

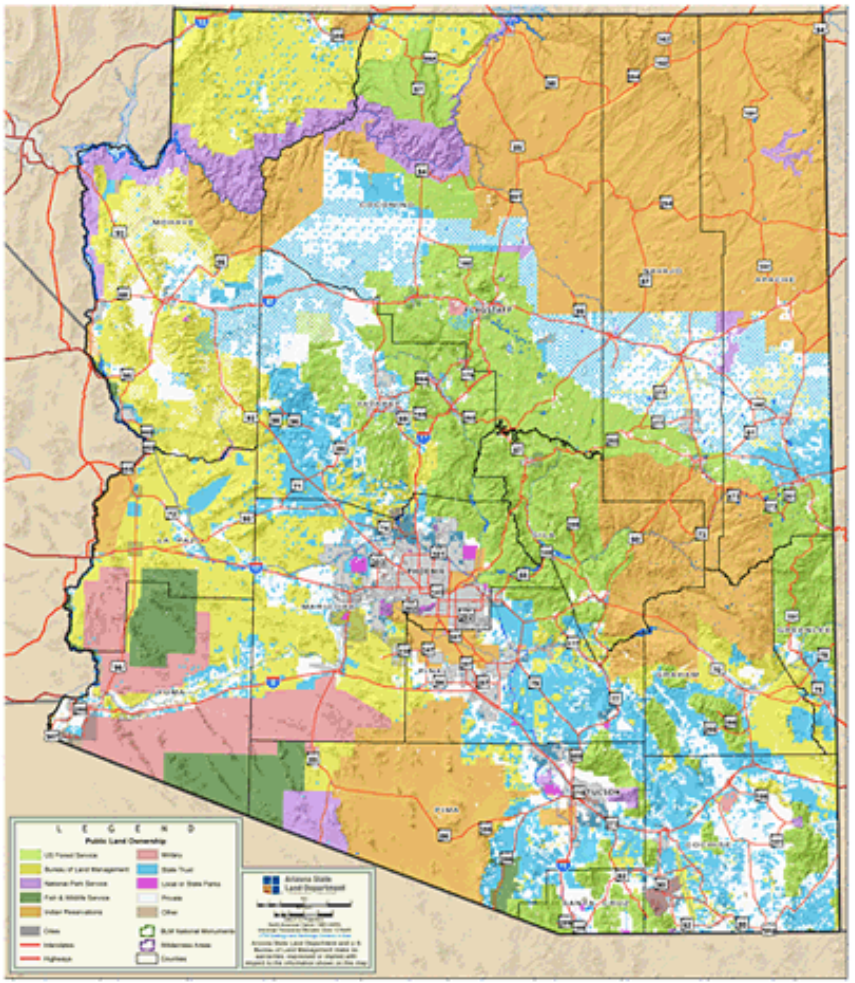
NATIONAL PARK SERVICE

# Land Use

Surface Management Responsibility

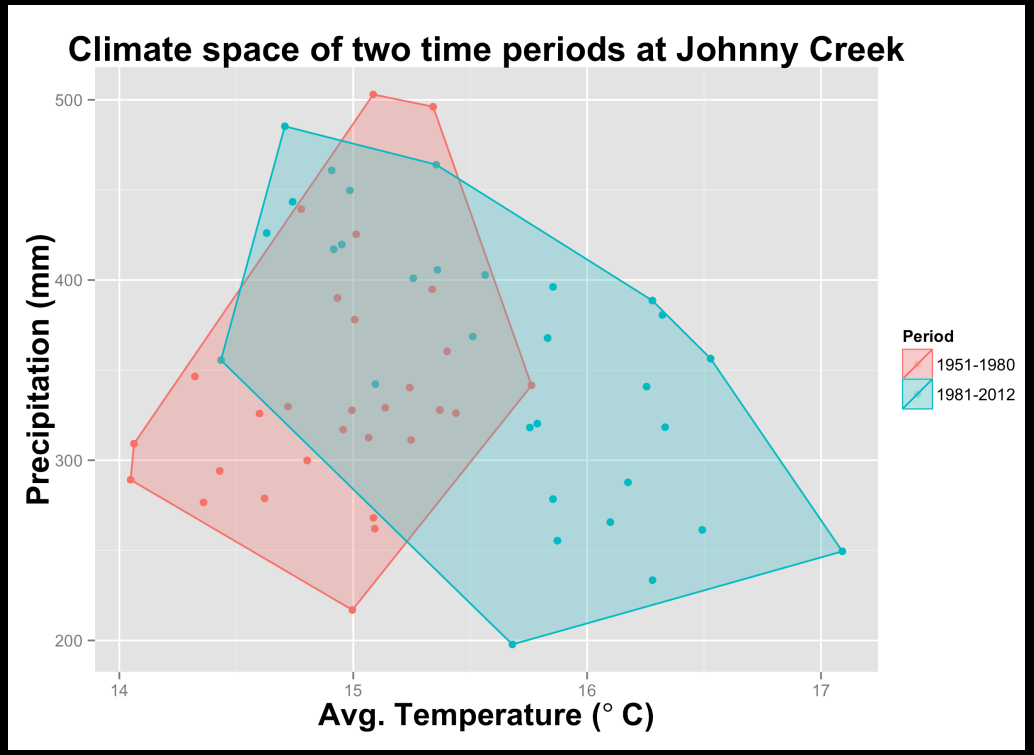
- Private
- State Trust
- BLM
- Forest Service
- Indian Reservation
- Military
- Wildlife Refuge
- Local or State Parks
- Other
- Natl. Parks
- Incorporated Cities
- County Boundary
- Interstates

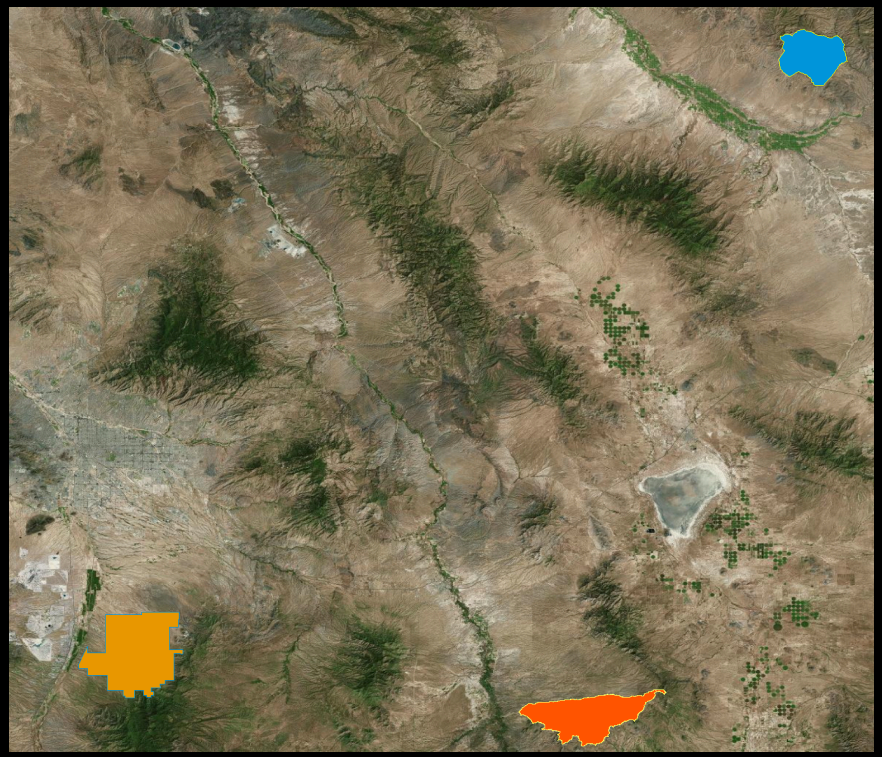
2011




## Land management

# Getting Warmer …and a Little Dryer?



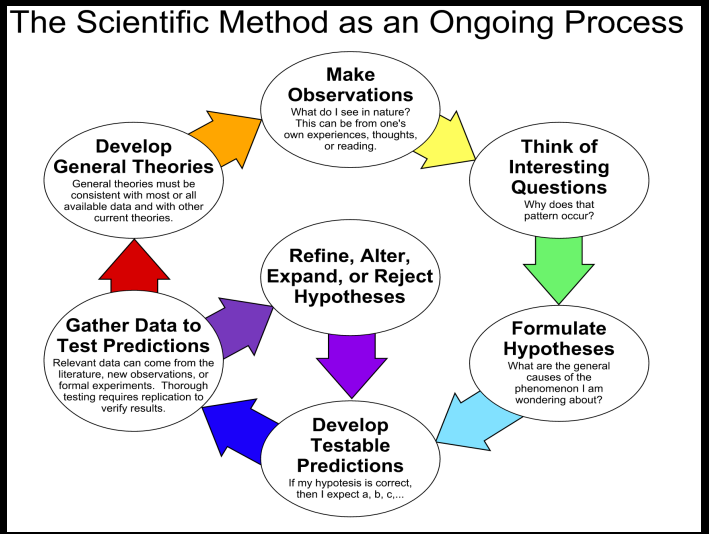
*PRISM Climate dataset (800m pixel size)*

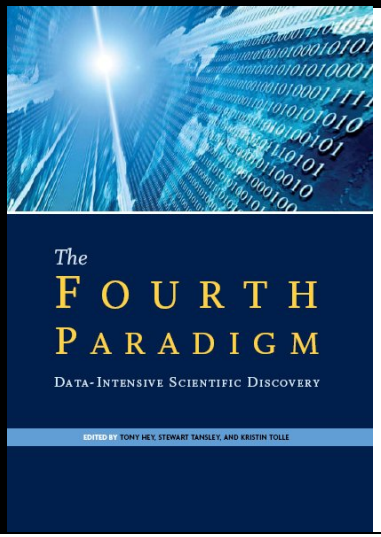
# Objectives

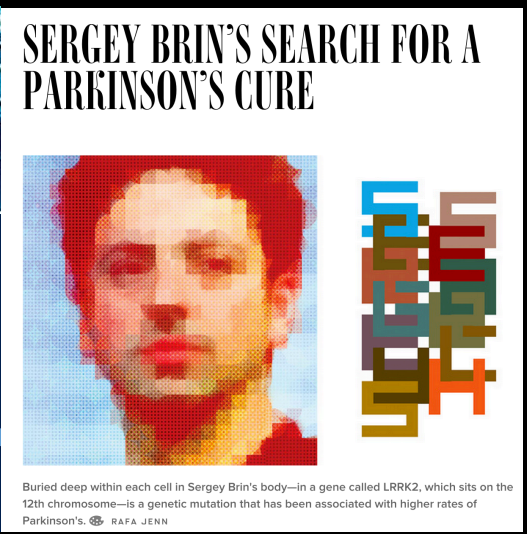
- *To investigate how climate and grazing activities affect grasslands vegetation productivity at a regional scale*

- *To develop tools to support NEPA (National Environmental Policy Act) planning by BLM*

# Exploring with Big Data




**Traditional Model**

1. Hypothesis: … GBA gene mutation might increase risk of Parkinson's.
2. Studies
3. Data aggregation
4. Analysis
5. Writing
6. Submission
7. Acceptance
8. Publication: *The paper notes that people with Parkinson's are 5.4 times more likely to carry the GBA mutation*

**Total time elapsed: 6 Years**

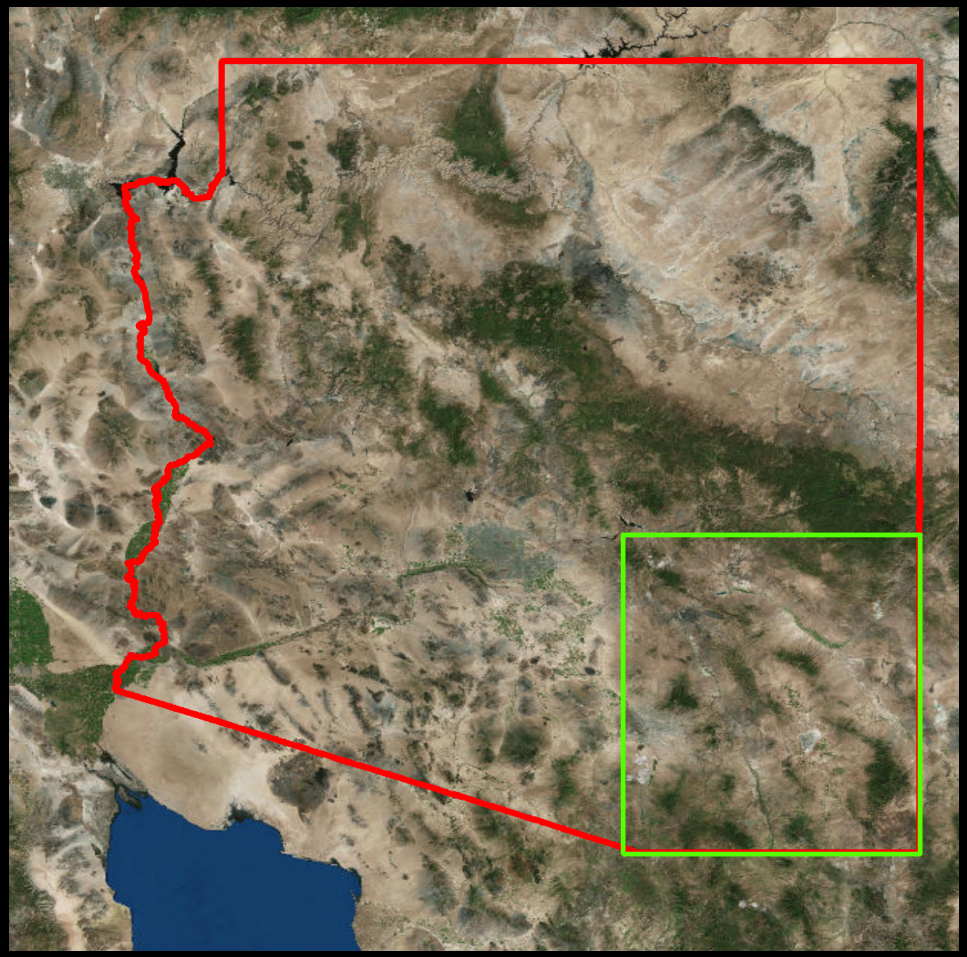
**Parkinson's Genetics Initiative**

1. Tool Construction
2. Recruitment
3. Data aggregation
4. Analysis
5. Presentation: *The results are reported at a Royal Society of Medicine meeting in London: People with GBA Mutation are 5 times more likely to have Parkinson's, which is squarely in line with the NEJM paper.*

**Total time elapsed: 8 Months**

*http://www.wired.com/2010/06/ff_sergeys_search*
*https://en.wikipedia.org/wiki/Scientific_method*

# Study Area



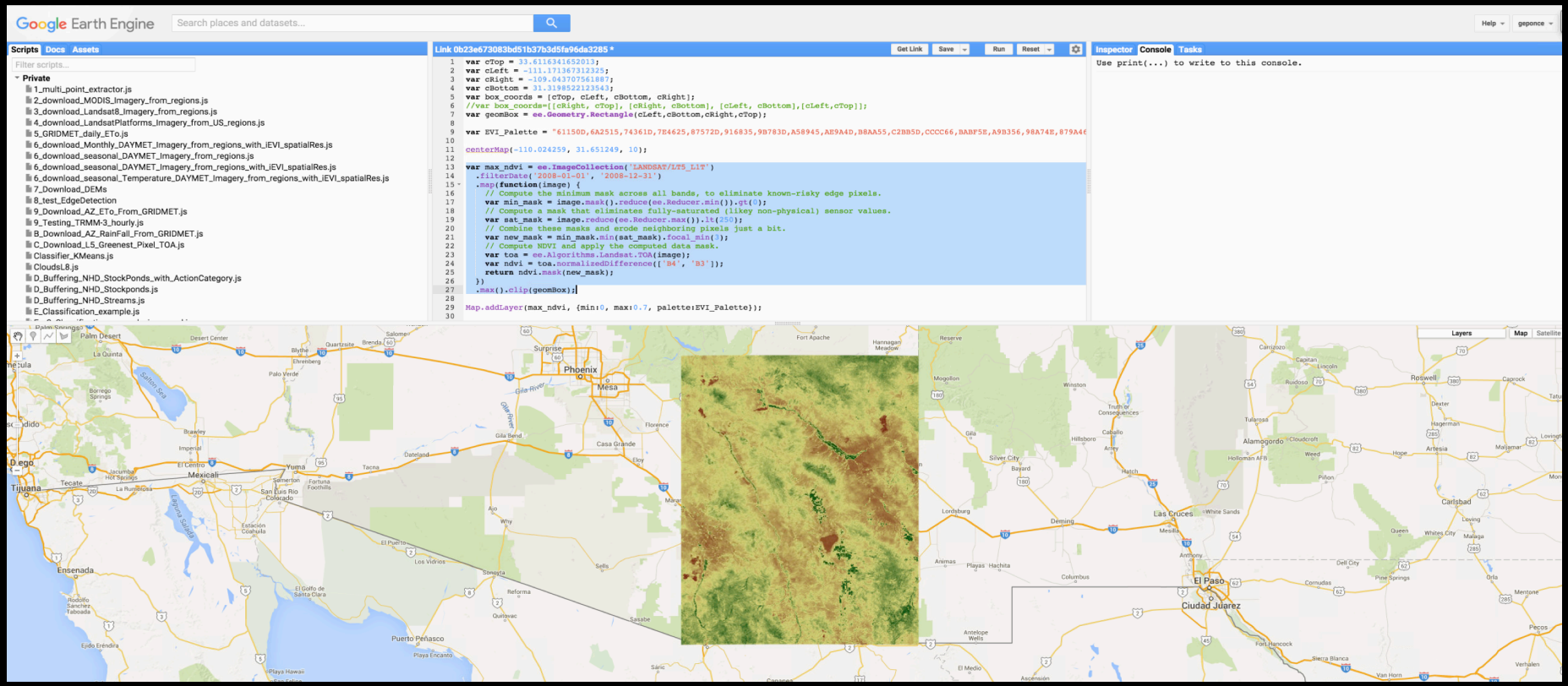
- *AZ Southeast corner*
- *Area ~ 48000 $Km^2$*
- *~ 67 Million Landsat Pixels (30m)*
- *Annual values (1998-2014) for a total of over 1 Billion observations (N)*
- *$N_{observations}$ X $P_{features}$*

# Data

**Dependent variable:**

- Annual Maximum NDVI from Landsat 5,7,8 as proxy for vegetation productivity

$$NDVI = \frac{\rho_{NIR} - \rho_{red}}{\rho_{NIR} + \rho_{red}}$$




Google Earth Engine

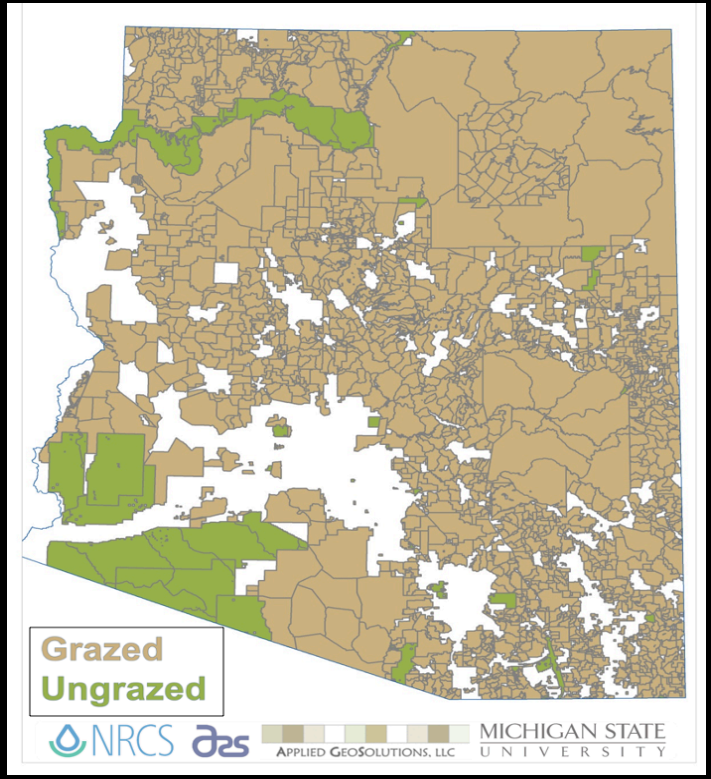
# …Data

**Predictors:**

- Climate data from GRIDMET and PRISM (SPEI & 33yr Normal)
- Topography information from USGS/NED 10m
- NHD (National Hydrography Dataset) layer for Streams
- MTBS (Monitoring Trends in Burn Severity) layer for fire events
- Solar radiation

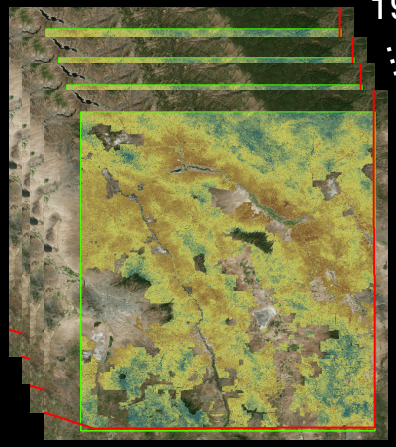
**Filtering options:**

- Land use layer --->
- Tiger-Roads database
- NHD for Ponds

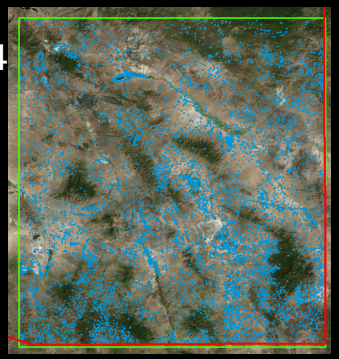


# Methods

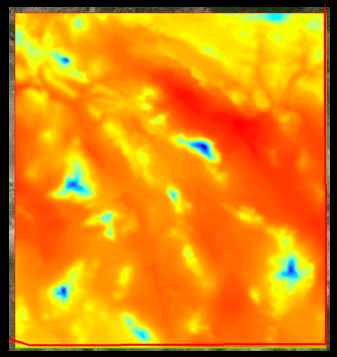
**Annual MaxNDVI 1998-2014**



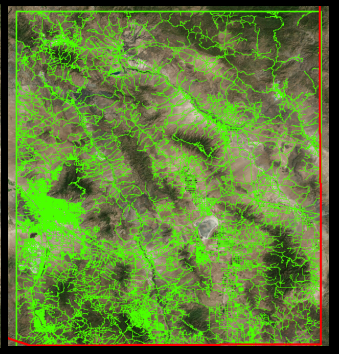

**Ponds**



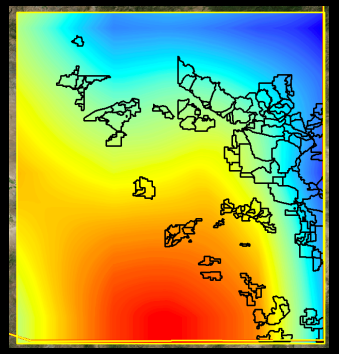
**33 Yr. Normal Precip.**



**Roads**



**SPEI**

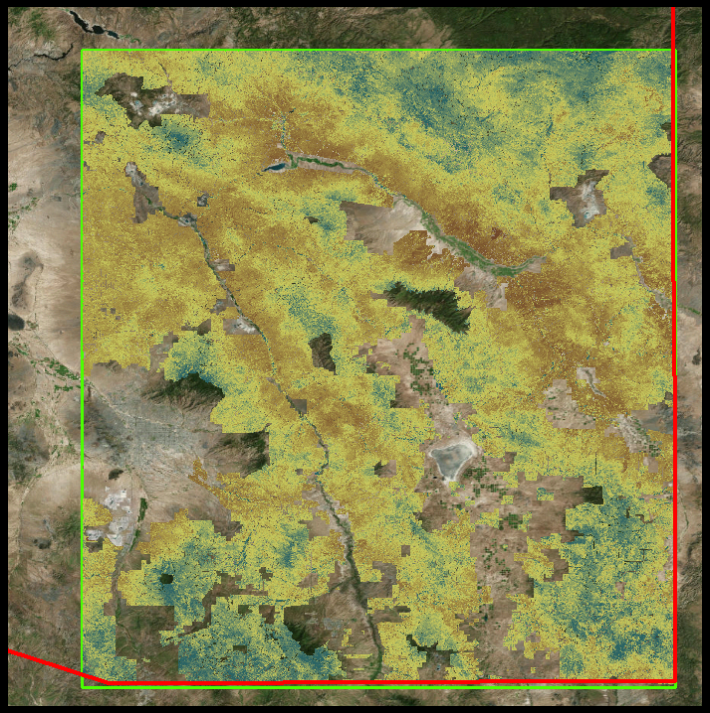


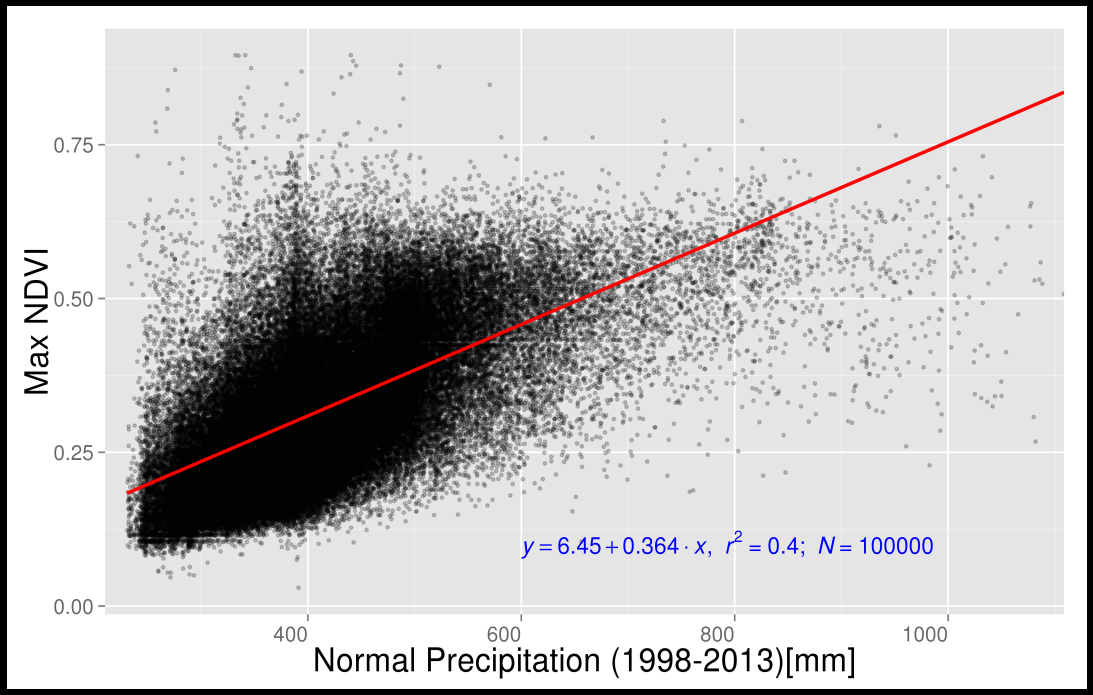
...

| | x | y | Year | Greenest | Normal800m | SPEI | Aspect | Slope | Landuse | Fires | SRad | Streams | Roads | Ponds | ANorth | ASouth | Landscape | FireEvent | PIXELID |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1: | -111.1712 | 31.31999 | 1998 | 0.4553448 | 469.5450 | 0.36221125 | 5 | 14 | NA | NA | 84755.55 | NA | NA | NA | NA | 1 | 3 | NA | 1 |
| 2: | -111.1712 | 31.31999 | 1999 | 0.5257357 | 469.5450 | 0.70527451 | 5 | 14 | NA | NA | 84755.55 | NA | NA | NA | NA | 1 | 3 | NA | 1 |
| 3: | -111.1712 | 31.31999 | 2000 | 0.3739657 | 469.5450 | -0.36683623 | 5 | 14 | NA | NA | 84755.55 | NA | NA | NA | NA | 1 | 3 | NA | 1 |
| 4: | -111.1712 | 31.31999 | 2001 | 0.4753914 | 469.5450 | -0.46150491 | 5 | 14 | NA | NA | 84755.55 | NA | NA | NA | NA | 1 | 3 | NA | 1 |
| 5: | -111.1712 | 31.31999 | 2002 | 0.4204004 | 469.5450 | -0.62695534 | 5 | 14 | NA | NA | 84755.55 | NA | NA | NA | NA | 1 | 3 | NA | 1 |
| --- | | | | | | | | | | | | | | | | | | | |
| 1141364356: | -109.0438 | 33.61150 | 2010 | 0.4351933 | 550.2344 | -0.59547744 | 4 | 5 | NA | NA | 61097.14 | NA | NA | NA | NA | 1 | 3 | NA | 67139080 |
| 1141364357: | -109.0438 | 33.61150 | 2011 | 0.4338821 | 550.2344 | -0.07974508 | 4 | 5 | NA | NA | 61097.14 | NA | NA | NA | NA | 1 | 3 | NA | 67139080 |
| 1141364358: | -109.0438 | 33.61150 | 2012 | 0.4148276 | 550.2344 | -0.35424048 | 4 | 5 | NA | NA | 61097.14 | NA | NA | NA | NA | 1 | 3 | NA | 67139080 |
| 1141364359: | -109.0438 | 33.61150 | 2013 | 0.5074590 | 550.2344 | 1.41141281 | 4 | 5 | NA | NA | 61097.14 | NA | NA | NA | NA | 1 | 3 | NA | 67139080 |
| 1141364360: | -109.0438 | 33.61150 | 2014 | 0.4480103 | 550.2344 | 0.57262203 | 4 | 5 | NA | NA | 61097.14 | NA | NA | NA | NA | 1 | 3 | NA | 67139080 |

By using the filtering options, only those pixels on grazed areas + off roads + off ponds buffer were used, we ended up with a *~ 600 Million observations*

# Linear fit



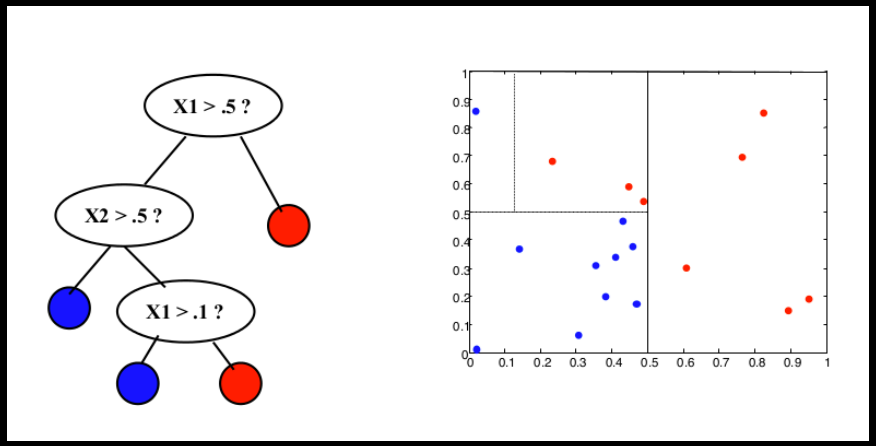


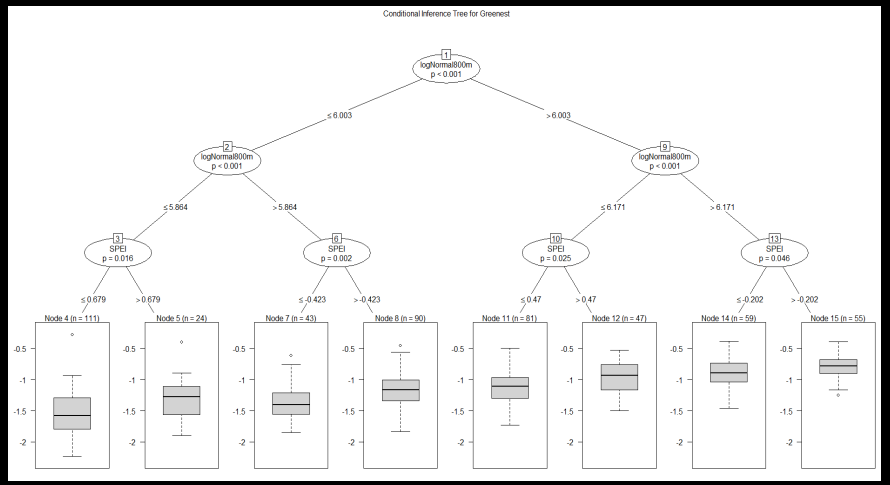

- For a multiple linear regression fit using all the predictors, the best $r^2$ was ~0.41
- Complex relationship as expected
- High variance

# Machine learning method: Random Forest

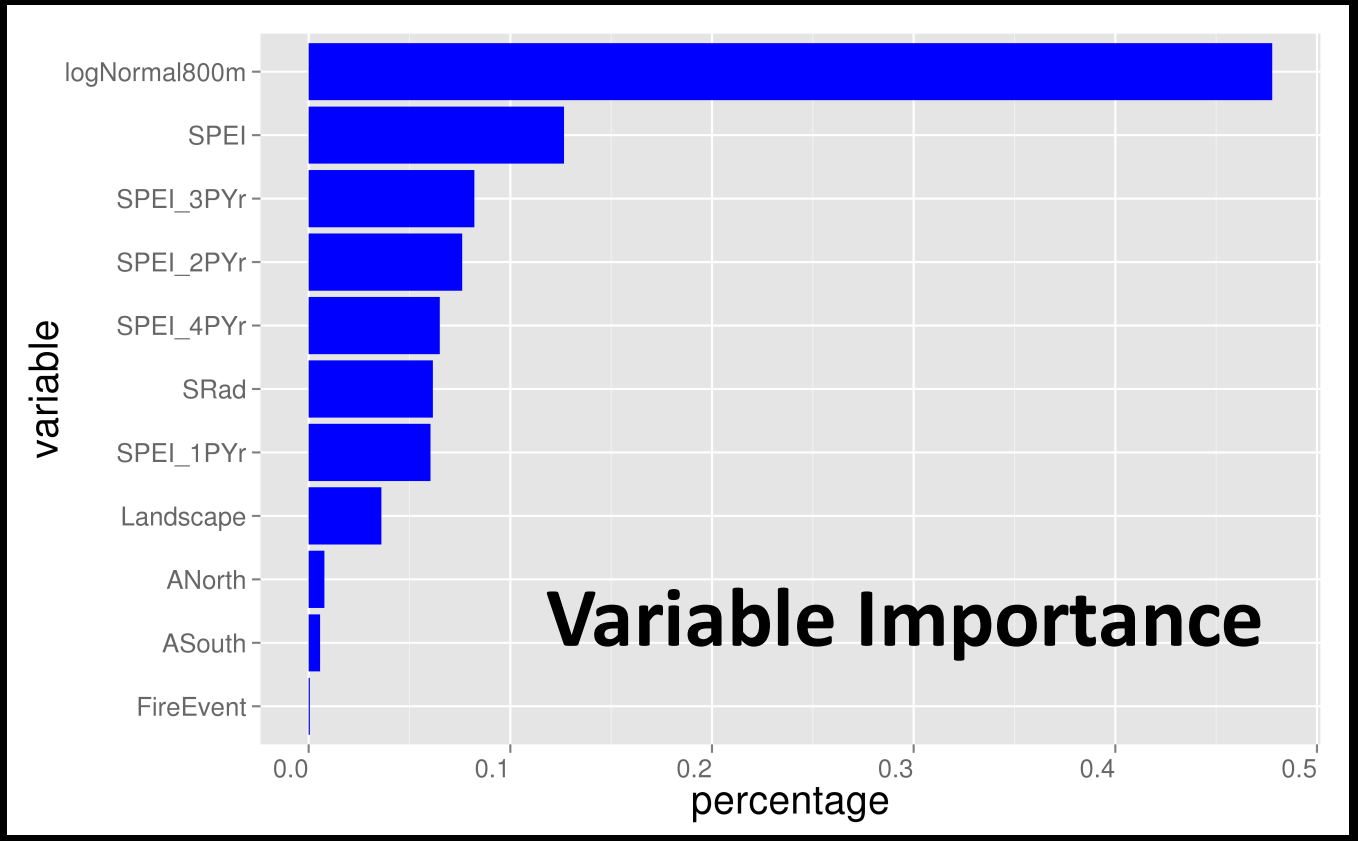
*(Breiman, 2001)*

- Decision tree based method
- Commonly applied for data classification, but can also be applied for regression
- Random forests are a way of averaging **multiple deep decision trees**, trained on different parts of the same training set, with the goal of **overcoming over-fitting problem** of individual decision tree
- By using our large data set, we extracted subsets for training and validation



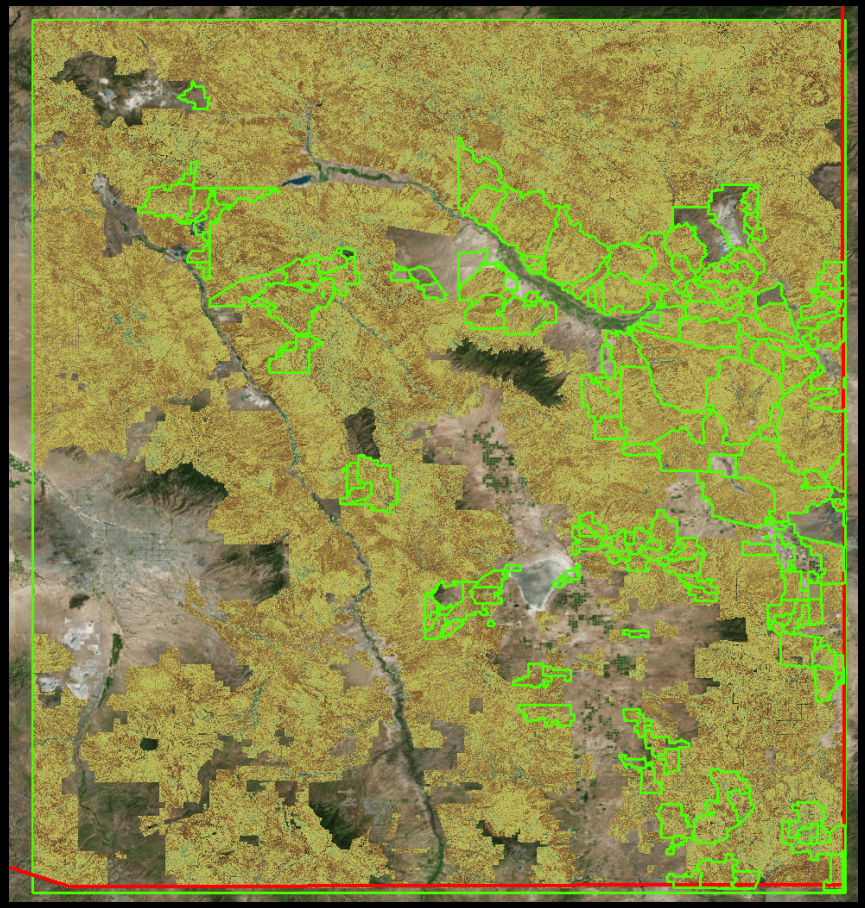
# Results



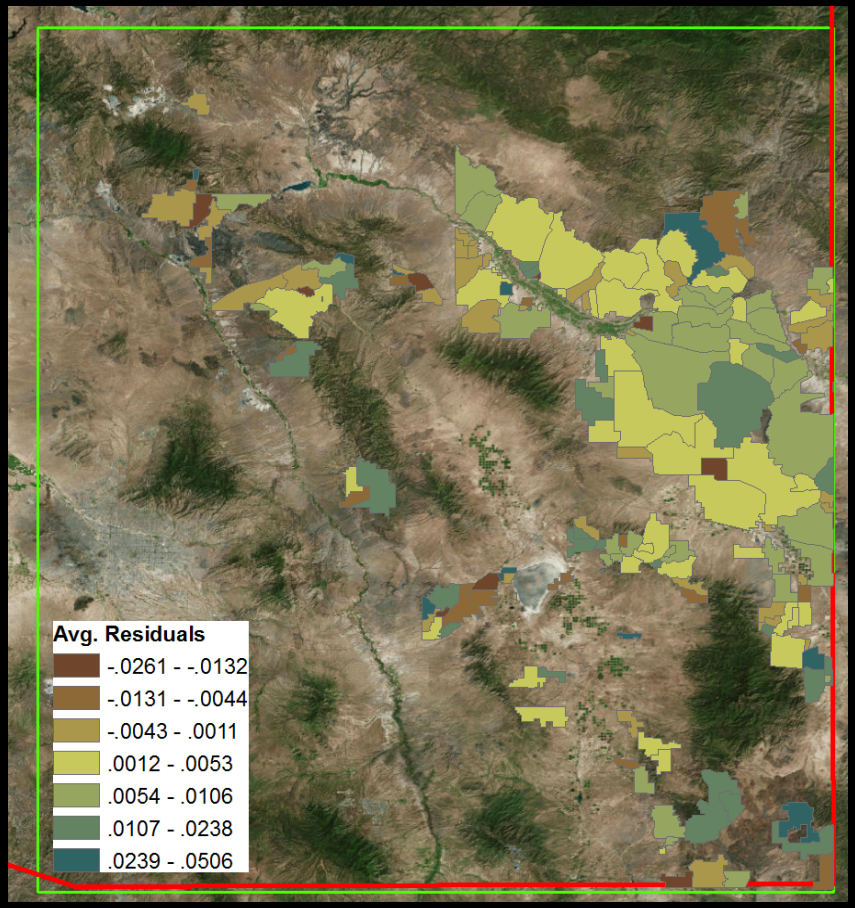

- Algorithm counts the number of times a split produce a lower RSS (Residual Sum of Squares)…

- Using the resulting random forest, a prediction was computed over the entire data set

# Results across study area…

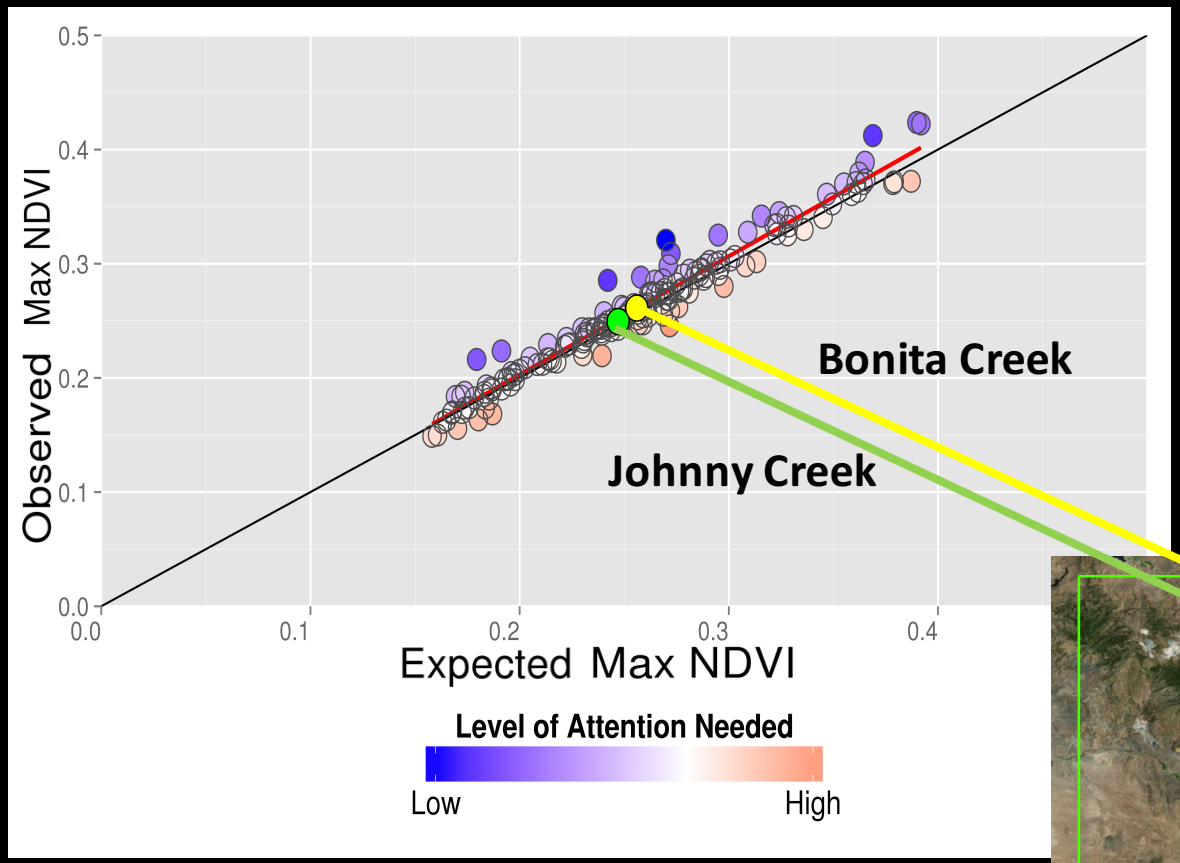


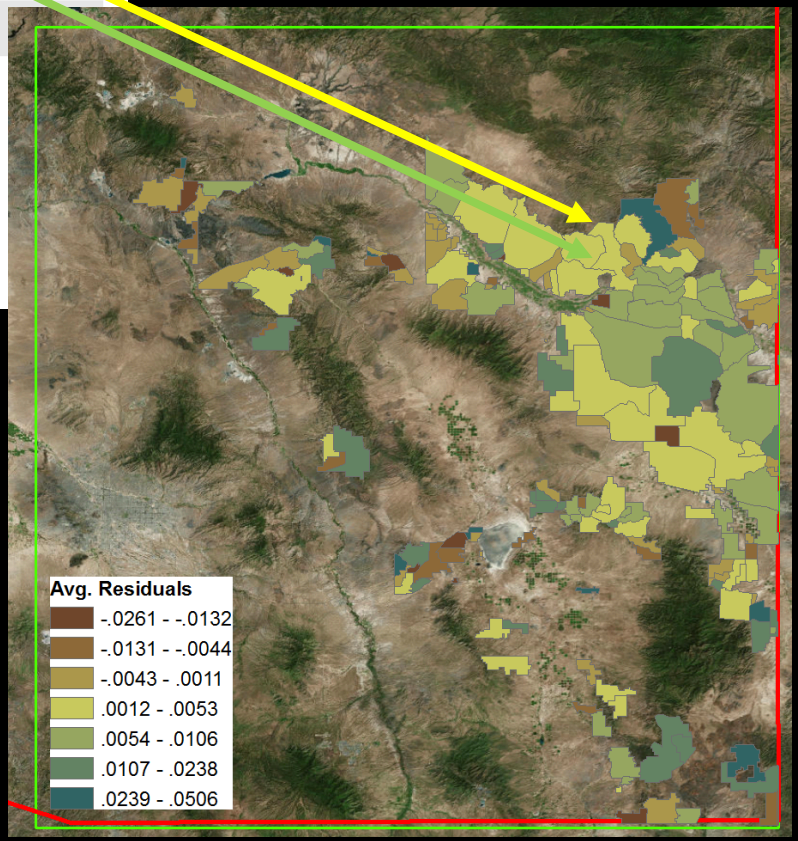
Avg. Residuals by Pixel




Avg. Residuals by spatial Unit (Safford Allotments)

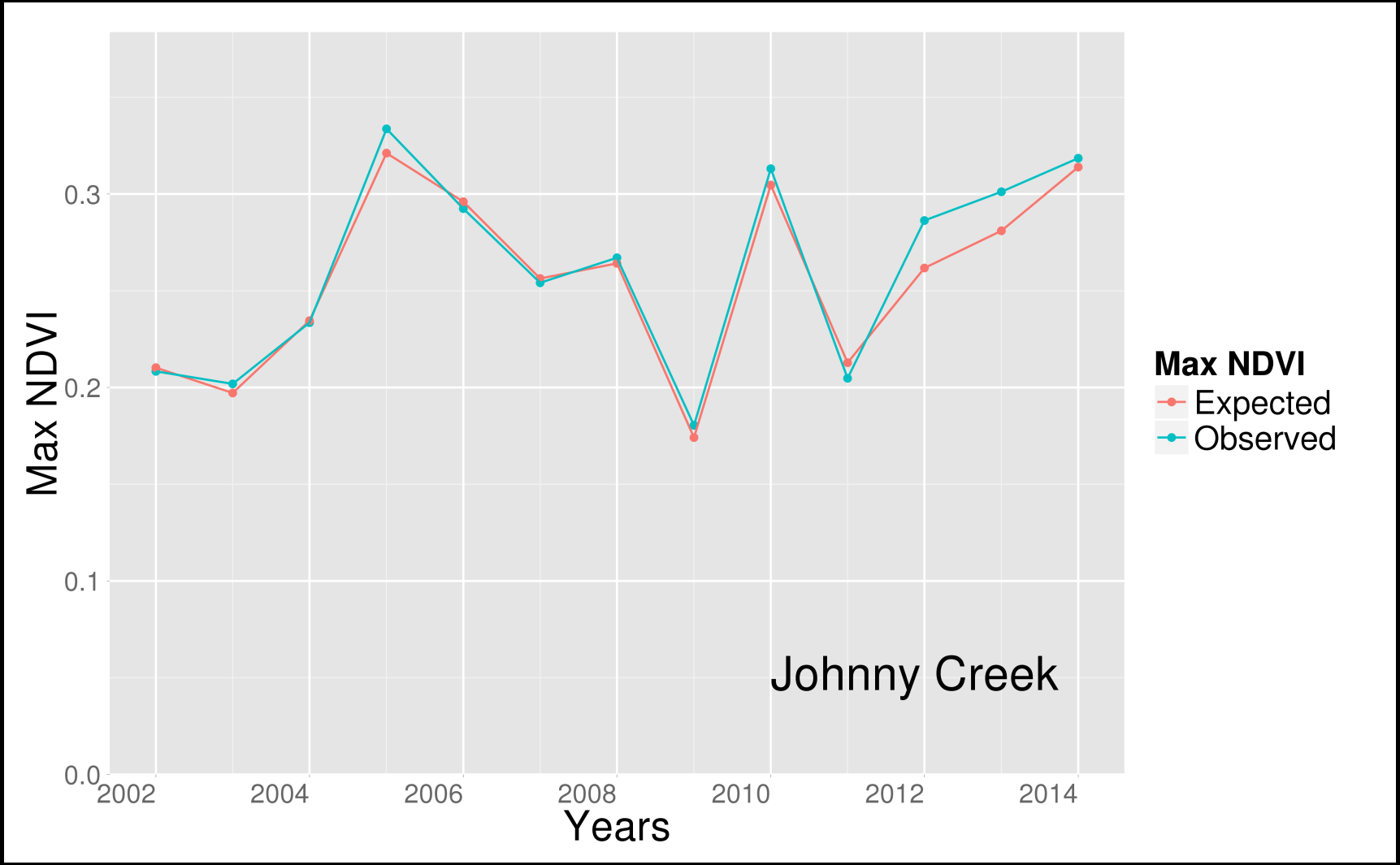
# Expected vs Observed by spatial unit



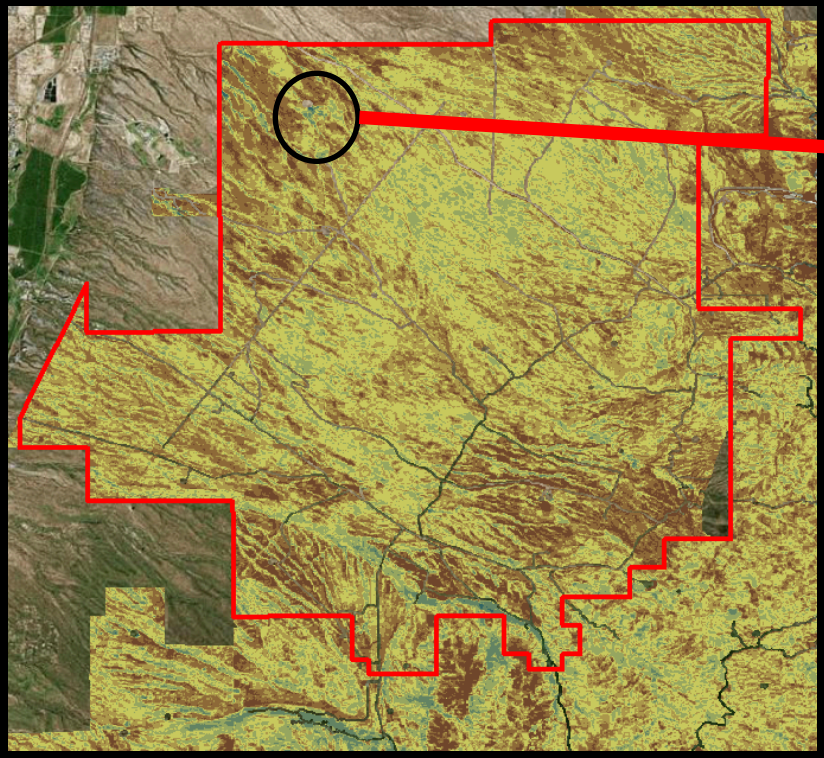



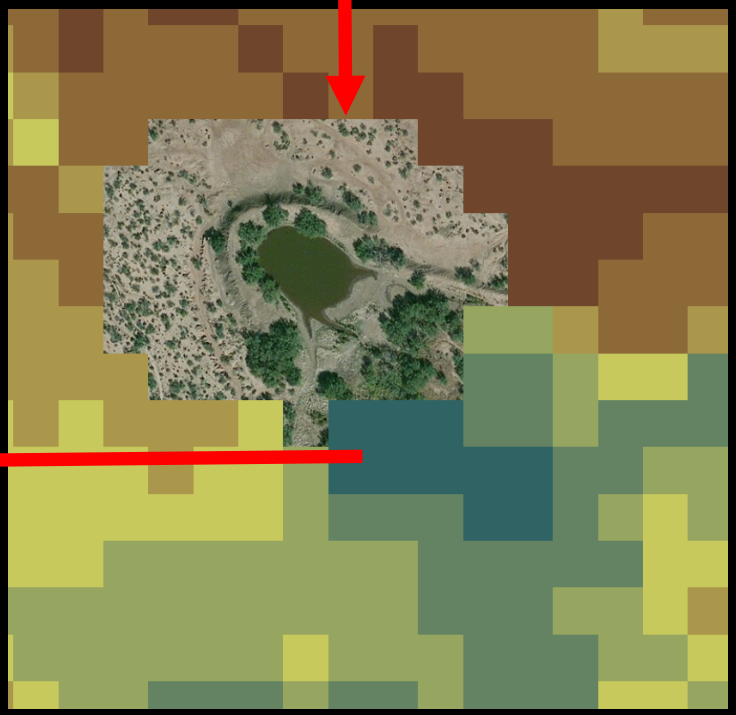

**By using this approach we can explore different spatial units in collaboration with agencies and field experts to support land assessment**

# Time series Exp. and Obs. by Allotment

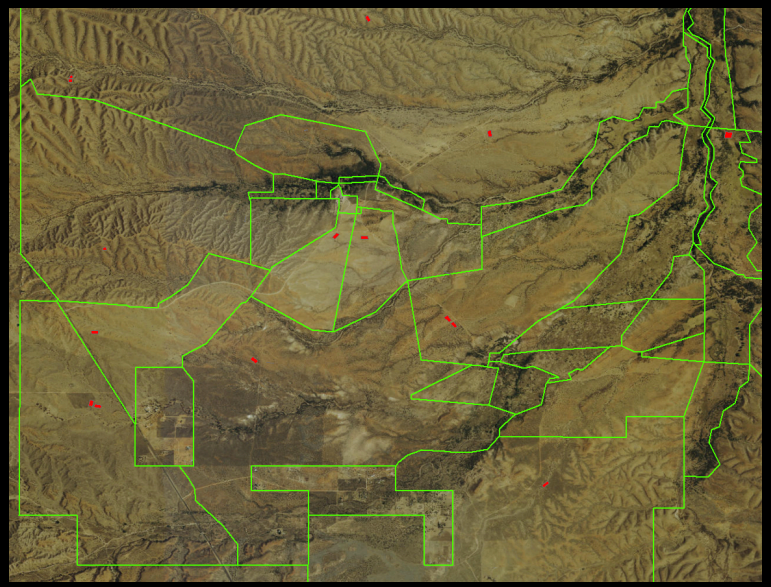


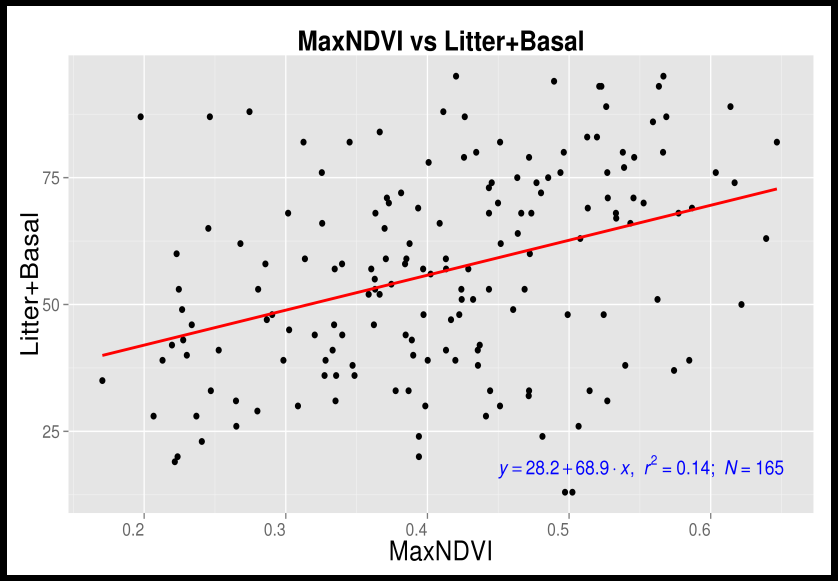
# SRER Average Residuals

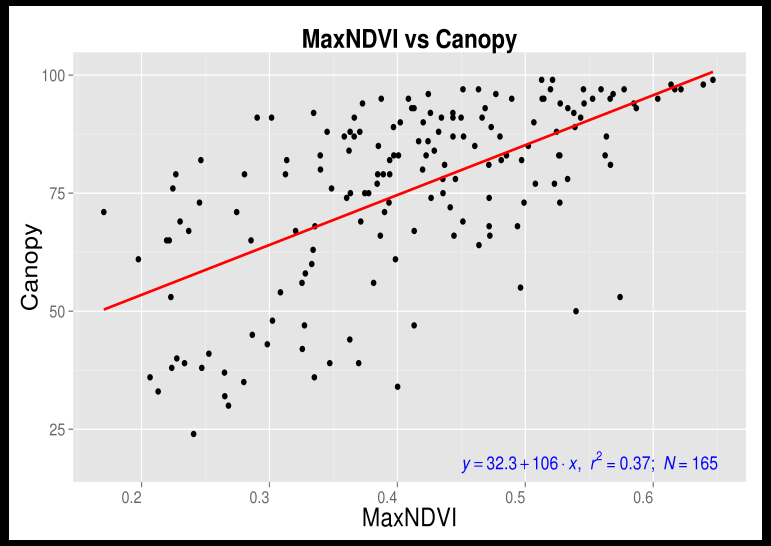


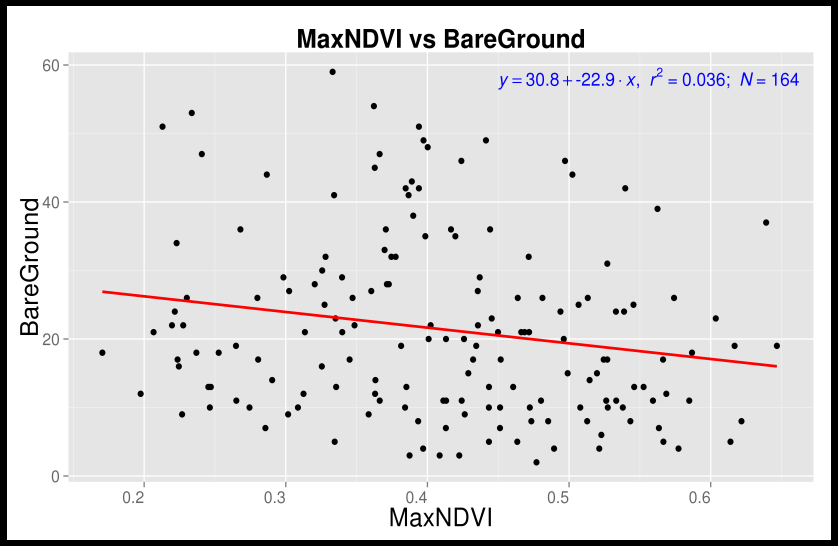


# Ground measurements at the Empire Ranch Area vs Max NDVI





# Concluding remarks

- *Regional scale approach to support land management operations*

- *Direct support for NEPA planning: NEPA imposes no requirements on the public. Rather, it directs Federal agencies to "utilize **a systematic, interdisciplinary approach** ... in planning and decision making which may have an impact on man's environment…"*

- *Aware of constraints (cloud issues, gridded climate,…), ground observations are limited, input data sets are not perfect*

- *Field review from experts will be critical*

- *Continue with an interdisciplinary approach will allow us to improve*

- *Within this workflow, we can plug-in other data sets as well as the application of different machine learning methods*

**Thanks…**

**Any Question?**